Lab01 – Numpy and Linear Algebra

CS3400 Machine Learning

Learning Outcomes

- Connect to the cluster and run a jupyter notebook interactive session
- Organize data into and interpret feature matrices
- Practice linear algebra operations with numpy

Overview

Welcome to CS3400: Machine Learning! This first lab is designed to help brush some cobwebs off that linear algebra you were forced to take and refamiliarize yourself with the tools we will be depending on in this class.

In this lab you will make sure you have access to the cluster and run an interactive Jupyter notebook session on ROSIE. In addition to the hardware available to you from ROSIE, using these interactive sessions will also ensure that you are working in a version-controlled environment that has the necessary software resources for you to complete our labs.

You will also be (re)introduced to the use and organization of Jupyter notebooks running a python 3 kernel. The exercises that you will perform were purposely created to pair with the first problem set due this week: Linear Algebra Review. In these exercises you will explore how we will be organizing our data in the class (numpy matrices), how we can select and manipulate entries in these constructs (slicing), and how the numpy library and its linalg (linear algebra) package can be leveraged to translate written linear algebra expressions into the computational domain.

Instructions

This lab is unique in that we are providing you with a Jupyter notebook stub. We have done this for multiple purposes:

1. To give you an organized start to completing the steps below.
2. To show you how your Jupyter notebooks should be organized for future labs.
3. Show some "tricks" of Jupyter notebooks that you might want to use in the future.

To complete this lab you will also need the linear algebra problem set that is due this first week.

Labs will typically be organized into separate experiments that you will be performing. In addition to experiments, although not for this lab (Lab 01), you will frequently be tasked with writing your own classes in python to import and use in your notebook.

In addition to the notebook stub we are also supplying a dataset that you will be able to read and manipulate in the notebook.

Jupyter Notebook – Running experiments

All Jupyter notebooks should follow the naming convention of <lastname>_lab01. The notebook should have a title, your name, and an introduction.

Each of the following steps (where applicable) should be performed in individual cells in your Jupyter notebook. Individual segments and their calculations should be in their own cells.

Experiment 1: Connecting to the Cluster

To perform experiment 1, you will need to follow the direction under experiment 1 in the provided Jupyter notebook. If you already have python 3 distribution and Jupyter notebook setup on your computer, you can open the provided .ipynb file. If you don't you can open the .pdf and follow the directions from there.

Experiment 2: Structuring Your Data

In this experiment you will need a running instance of Jupyter notebook. You should now open the provided .ipynb file and start following the directions under experiment 2!

Experiment 3: Linear Algebra in Numpy

Complete experiment 3 in the same notebook. Follow along in the stub you were provided.

Bonus Material: Additional Indexing Topics

At the end of the notebook there is an additional section going over some more indexing/slicing methods. You should take a look at this material. It may help in this course.

Submission Instructions and Grading Criteria

Save the notebook as a PDF named lastname_lab01.pdf. Upload the PDF through Canvas. Print off and submit a hard copy of this pdf during the following week's lab meeting.

I will be looking for the following:

- A title, your name, an introduction (including your own summary of the lab), and your answers to the reflection questions at the top of the notebook in Markdown.
- Proper effort was put into each part of each experiment
- That you have the appropriate number of plots and they look reasonable. I will be checking for proper axis labels.

Presentation	10%
Followed submission and formatting instructions	5%
Plots: axes are properly labeled, used correct axes for variables, points were colored as required, lines were coloring as required, used a legend, chose appropriate axes limits to make plot readable and do not cause misleading interpretations, etc.	5%
Experiment #1: Connecting to the Cluster	**15%**
Experiment #2: Structuring Your Data	**40%**
Load Data	10%
Plots	10%
Slicing / Indexing Operations	10%
Summary Statistics	10%
Experiment #3: Linear Algebra in Numpy	**30%**
Exercises 1 – 10	20%
Linear Regression Predictions	10%
Exceeds Expectations	**5%**